



Electrical signal integrity considerations for HP BladeSystem

technology brief

Introduction.....	2
What is signal integrity	2
Challenges.....	4
Significant factors affecting signal integrity	6
Dielectric losses	6
Skin effect	6
Impedance discontinuities	6
Stubs	7
Crosstalk	8
Design goals	9
Target fabrics	10
Infrastructure architecture	10
Implementation.....	11
Specification library	12
Board trace lengths	12
Board layout and materials	12
Summary	15
For more information.....	16
Call to action	16

Introduction

A requirement for any server architecture is that it continues to meet future customer needs by supporting a built-in capability to provide sustained high performance as technology driven bandwidth enhancements become available.

The BladeSystem c-Class enclosures (c3000, c7000) are architected to ensure that they can support upcoming technologies and increasing demand for bandwidth and power for at least 5 to 7 years. This technology brief discusses the effort to design, model, standardize, and document the electrical requirements for the high-speed interfaces of the HP BladeSystem. This applies to all HP BladeSystem switch modules, mezzanine cards, server blades, and midplanes that use the high-speed interface.

What is signal integrity

Signal integrity engineering ensures that electronic circuits communicate reliably in the presence of channel impairments. Over short distances and at low bit rates, a simple conductor can easily transmit a signal from point A to point B with good fidelity (Figure 1).

Figure 1. Lassy signal line depiction over short distances and low bit rates



However, at high bit rates and over longer distances, various effects can degrade the electrical signal to the point where data transmission errors occur, resulting in a system malfunction (Figure 2).

Figure 2. Lassy signal line depiction over longer distances and higher bit rates



As signal speeds increase additional effects inherent in the transmission media environment emerge as inhibitors to successful signal delivery. These effects include:

- the physical characteristics of the material used for transmission, including materials adjoining the transmission media
- the interference caused by the interaction of multiple simultaneous signals
- the noise introduced at connection points as the signal passes from one component to another

While in the 2002 (and earlier) timeframe, effective signal analysis calculations required using Ohm's Law, but in the 2008 (and beyond) timeframe these calculations require a more sophisticated approach using Maxwell's Equations. Figure 3 depicts how fundamental design considerations have increased in complexity as transmission speed has increased. To better manage this increased complexity, a dedicated signal integrity research group is tasked to create, develop, and publish design guidelines. These guidelines illustrate design implementation techniques necessary to avoid the pitfalls created by increased transmission speed. These guidelines are then made available to the product design specialists to apply as necessary. Strict adherence to these guidelines is required.

Figure 3. This figure contrasts signal integrity analysis in 2008 versus 2002 (and earlier)

2002	2008
$V = I R$ (Ohm's Law)	$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$ $\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}$ $\nabla \cdot \mathbf{D} = \rho$ $\nabla \cdot \mathbf{B} = 0$ (Maxwell's Equations)

Challenges

The NonStop signal midplane in a BladeSystem enclosure is capable of conducting extremely high signal rates of up to 10 Gb/s. Each half-height server blade has the cross-sectional bandwidth to conduct up to 160 Gb/s per direction. In a c7000 enclosure fully configured with 16 half-height server blades, the aggregate bandwidth is up to 5 Terabits/sec across the NonStop signal midplane. This is bandwidth between the device bays and the interconnect bays only. It does not include traffic between interconnect modules or blade-to-blade connections.

Significant signal losses occur as a function of increased speed, distance, and discontinuities, but the loss can be mitigated in several ways:

- Improving the passive signal channel to reduce frequency-dependent losses
- Improving the driver and receiver silicon to compensate for frequency-dependent losses in the passive signal channel
- Invoking a combination of these two approaches

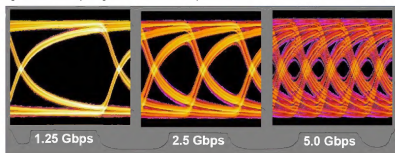
There are a few main challenges to signal integrity:

- Insufficient line bandwidth, which results in attenuation and inter-symbol interference
- Impedance discontinuities at connectors which results in reflection and therefore signal distortion
- Cross-talk which reduces signal quality

As the signal frequency increases, many factors begin to affect the signal's integrity. An effective way to view the potential integrity of the signal is to use an eye diagram. The eye diagram is a collection of timing-synchronized overlaid oscilloscope traces depicting a large set of differing bit sequences. The bit sequences that make up the transmitted signal create unique energy signatures that affect the signal's wave shape, making the signal sequence dependent. In other words, the order of 1's and 0's transmitted can impact the integrity of the transmission. The "eye" in the eye diagram is shown in Figure 4; it is the opening in the center of each example. This opening depicts the edge transitions and the time available for the receiver circuit to determine the value of the received signal. Follow a trace in Figure 4 left to right from one eye to the next. If the trace was high in one eye that trace will reach a higher level in the next eye. If the trace was low in one eye it does not reach as high a level in the next eye. This is called inter-symbol interference.

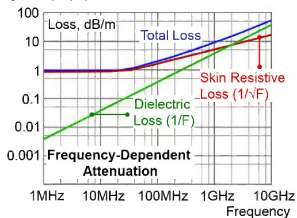
Figure 4 illustrates eye closure due to attenuation and bandwidth limitations, these create inter-symbol interference, eventually limiting the receiver's ability to distinguish one bit from another. Noise causes variation up and down on each trace in the eye diagram. Jitter is uncertainty in the horizontal time where a trace goes from high to low or low to high. These degradations close down the available headroom in the eye opening.

Figure 4. Measured eye diagrams from a 40" PCI Compliance ISI trace



Increasing frequency significantly increases attenuation. Skin resistive loss and dielectric loss are the primary components of frequency-dependent attenuation, as shown in Figure 5.

Figure 5. Frequency dependent attenuation



Significant factors affecting signal integrity

A transmitted signal arriving at the receiver must maintain sufficient magnitude and quality to be reliably recognized. As signal frequency increases, previously inconsequential factors emerge as new obstacles.

Line bandwidth is affected by attenuation introduced by factors such as dielectric loss and the skin effect. Additionally, the reflections produced at impedance discontinuities, specifically at connector interfaces, become significant because they generate noticeable signal distortion. At these speeds, crosstalk between transmission lines becomes a limiting factor. These factors are discussed briefly below.

Dielectric losses

Dielectric loss is caused by the electromagnetic field associated with a signal traveling along a conductor and its interaction with the adjacent material, in this case the printed circuit board. This interaction contributes to high frequency signal attenuation. Dielectric loss is a principal cause responsible for frequency-dependent losses within printed circuit boards.

Skin effect

The skin effect essentially increases the resistance of a conductor as the signal frequency increases. As the signal frequency increases, the signal current is forced toward the surface of the conductor by the magnetic field and to vacate the center of the conductor. The signal current is said to travel at the "skin" of the conductor. This has the effect of limiting the accessible area of the conductor that is available to carry the signal current, increasing the resistance of that conductor.

Impedance discontinuities

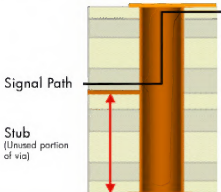
Potentially, impedance discontinuities may occur at any point where the transmitted signal is transferred from one component to another. This transfer normally represents a physical structure change in the signal path, such as, connectors, plated through-holes for connector pins, vias for transitioning printed circuit board (PCB) layers, and differing board materials. At the point of the discontinuity, an impedance change creates reflected energy back to the source of the signal. This becomes a new signal that adds to the desired signal at an incorrect time effectively reducing the quality of the transmitted signal.

Stubs

In PCB design, through-hole vias connect the signal path from one layer to another. As shown in Figure 6, many instances of this path leave a portion of the via unused. As signal speed increases, this unused portion of the via becomes a transmission line that can create significant reflections that can seriously reduce signal quality. Any conductor that extends off the main conductor some length and then stops is called a stub.

A through-hole via connecting the signal path from one layer to another that extends beyond the main path is an example of a stub. As shown in Figure 6, traces changing layers frequently leave such an unused portion of the via. A portion of the signal traveling on the trace will go down the stub and reflect off the end and return to the trace after the delay of going down and back. This reflected signal is added to the traveling signal. At high enough data rates this delay is long enough compared to the bit length to adversely affect a portion of the next bit. This distorts the signal waveform.

Figure 6. Cross sectional view of a plated through hole, the portion of the via that is not used for the signal path is known as a stub



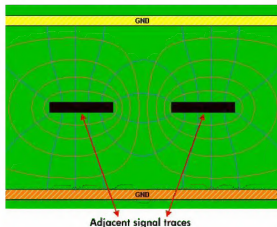
Crosstalk

As current moves through a conductor it creates an electromagnetic field. When two (or more) conductors run parallel to each other, the inductive and capacitive coupling between the paths can lead to interference. This interference is also known as crosstalk. Essentially, crosstalk is the coupling of a signal from one conductor into another adjacent conductor and is exacerbated by increasing transmitted signal speed and reducing the distance between adjacent conductors.

Crosstalk results in adding a portion of a signal to its adjacent signals. The externally generated portion of any signal is noise to the original signal. Noise reduces the eye opening and the probability of correctly transferring data on that conductor. Figure 7 depicts the magnetic field that couples two traces and the fact that this coupling diminishes with distance. To the extent possible, the conductors in the board are routed with appropriate spacing to minimize crosstalk.

Crosstalk also occurs between the vias in the pin fields of the connectors and in the connectors themselves. These are complex three dimensional structures with non-uniform electric and magnetic fields.

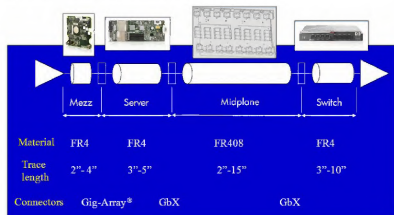
Figure 7. Cross sectional view of parallel signal traces depicting electromagnetic fields and their interactions



Design goals

The BladeSystem c-Class uses a high-speed signal midplane that provides the flexibility to intermingle server blades and interconnect fabrics in many ways to solve a multitude of application needs. The NonStop signal midplane is unique because it can use the same physical traces to transmit GbE, Fibre Channel, 10 GbE, InfiniBand, Serial Attached Technology (SAS), or PCI Express signals. As a result, customers can fill the interconnect bays with a variety of interconnect modules depending on their needs. Figure 8 depicts the signal path across the BladeSystem architecture. It is organized by module with divisions shown at connection points and indicates the board material, trace length range, and connector type for each element. The variance in the trace length parameter listed in Figure 8 is due to the size of the circuit board; the larger the circuit board the longer the traces may be.

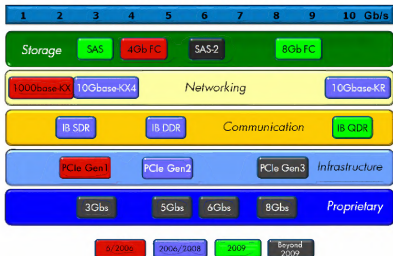
Figure 8. Signal path across the BladeSystem architecture



The NonStop signal midplane can transmit signals from different I/O fabrics because of similarities in the physical layer of those fabrics. Serialized I/O protocols such as GbE, Fibre Channel, 10GbE, SAS, PCI Express, 1000Base-KX, 10GBase-KX4, and InfiniBand along with the 8 Gb Fibre Channel are all based on a physical layer that uses two or more differential pairs with a serializer/deserializer interface.

The BladeSystem c-Class enclosure was designed and built to ensure that it can support upcoming technologies and their demand for bandwidth and power for at least five to seven years. This required looking in the infrastructure long before certain fabrics would be available to be tested (Figure 9).

Figure 9. Fabric technology targets



Additionally, the design had to incur minimal additional cost for infrastructure to support fabrics yet to be developed, leverage industry standard components, and interoperate with the dozens of companies and multiple divisions involved.

Over its lifecycle, the BladeSystem c-Class architecture must support dozens of server blades, more than 50 Mezzanine Cards (MEZZ) and I/O modules, and multiple enclosure designs.

Target fabrics

The HP BladeSystem provides high-speed interfaces between server blades and switch modules. Server blades may contain components connected to the high-speed interfaces on their motherboards or on mezzanine cards. The main high-speed interface is expected to be used for connecting server blades to switch modules with high-speed differential signals. In various implementations, these channels are used for routing Ethernet, Fibre Channel, or other types of signaling.

Infrastructure architecture

For single-wide switch modules, each main high-speed interface consists of a group of four pairs of signals, available for use as two lanes (two transmit pairs and two receive pairs) connected to a server blade. Additionally, for single-wide switch modules, there is also an inter-switch link available that allows two adjacent switch modules to communicate for failover, connectivity, and so forth. It consists of eight pairs of signals available for use as four lanes (four transmit pairs and four receive pairs).

For double-wide switch modules, the main high-speed interface consists of two groups (eight pairs) of signals that are available from each server blade. These are used to route signals that require up to four lanes, such as 10 Gb Ethernet using the 3.125 Gbaud SERDES¹ type of interface.

Implementation

In order to guarantee success and manage costs effectively, the BladeSystem design team separated the system into its component parts, illustrated in Figure 8, created and applied significant modeling/simulation techniques, and then verified, through empirical data, that the models accurately reflected the physical system. The team then created an extensive documentation/specification library and implemented a design review process that ensures continued adherence and success. Specific steps included the following:

- Assisting the internal product teams in signal integrity best practice; providing layout/routing guidelines; including optimizations for 10Gbs
- Assisting in the design and validation of top tier mezzanine and switch designs; assisting vendors with signal integrity related direction and in-house hardware measurements
- Developing accurate channel models
- Ensuring that the NonStop midplane can handle up to 10Gbs signaling
- Developing signal integrity tools and models to help develop lower tier independent hardware vendor (IHV) solutions

Achieving this level of bandwidth between bays required special attention to maintain the signal integrity of the high-speed signals. HP took three key steps to maintain signal integrity:

- Using general best practices for signal integrity to minimize end-to-end signal losses across the signal midplane
- Moving the power into an entirely separate backplane to independently optimize the signal midplane
- Providing means to set optimal signal waveform shapes in the transmitters, depending on the topology of the end-to-end signal channel

Following best practices for signal integrity was important to ensure high-speed connectivity among all server blades and interconnect modules. To aid in the design of the signal midplane, HP involved the same signal integrity experts that design the HP Superdome computers. Specifically, HP paid special attention to several best practices:

- Controlling the differential impedance along each end-to-end channel on the PCBs and through the connector stages
- Planning signal pin assignments so that receive signal pins are grouped together yet isolated by a ground plane from the transmit signal pins in order to minimize crosstalk
- Keeping the overall channel short to minimize losses
- Increasing the length of short traces in order to minimize reflections
- Routing signals in groups to minimize signal skew
- Reducing the number of through-hole via stubs by carefully selecting the layers to route the traces, controlling the PCB thickness, and back-drilling long via-hole stubs to minimize signal reflections

¹ The 3.125 Gbaud SERDES is a serializer/deserializer (SerDes) data interface used to transmit data between chips.

Specification library

All server blade and midplane channels (and segments of channels) are required to meet the electrical specifications developed by HP, allowing separate vendors to create interoperable parts of the channel. The server blade vendor must ensure that any channel terminating at an IC on the server blade meets the same requirements as the combined effects of the server blade-to-MEZZ and MEZZ channel segments. The following sections describe examples of generic electrical characteristics of the high-speed interfaces between the switch module and the server blade or MEZZ, as well as the inter-switch link.

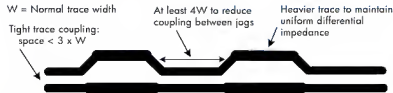
Board trace lengths

If the signal terminates in an IC, the trace length includes the electrical length of any series capacitors and their vias and any via needed to connect to the IC. If the design requires other external components (for instance, terminating resistors), the trace length also includes any electrical length necessary for those components. GbX[®] connectors have different electrical lengths between the contacts in different rows. This difference between rows must be compensated by a trace length difference for signals that make up a differential pair. Signals must include additional trace length for each GbX connector. The extra trace must be added as close as possible to the GbX connector to avoid signal distortion at very high frequencies caused by differential trace coupling. Compensation for this difference is made on the midplane so server blade and switch module vendors need only match the traces of a pair.

The differential pairs are classified as "tightly" coupled when the traces are less than 3x width apart. This maintains the differential character of the signals. The section of the differential pair that is spread apart is given a slightly larger line width to maintain the differential impedance as exactly as possible (see Figure 10). Trace length matching should be done as close as possible to the point where the length difference occurs.

Ground layers must fully surround the connector pins for the GbX connector on the server blade; failure to do this results in a mismatched impedance in the connector field and adds skew and common mode noise to the signals. Given the proximity of the GbX with the edge of the server blade PCB, the ground layers may need to be exposed to the PCB edge.

Figure 10. Jog layout for length matching with tight coupling



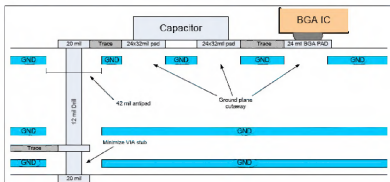
Board layout and materials

High-speed signal routing (>2Gbit/sec) is extremely critical and should follow all of these guidelines, or ascertain through testing and simulations that it meets the required functionality.

¹ The GbX connector is a high-density differential connector providing increased density, impedance matching, and low crosstalk.

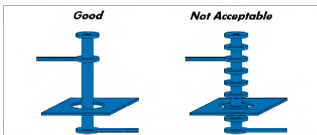
The impedance of channels that include components and component pads should be adjusted by removing some of the adjacent ground plane under these components and pads. To prevent crosstalk and impedance discontinuities in other signals, no signals are routed under these "cutaways" in the ground layer. Vias should use the smallest practical drill size and only be used to connect to components or connectors. Via stubs should be limited in size. In cases where the traces must be routed on layers with stubs longer than specified, the vias should be back-drilled to reduce the stub length, as previously described. Vias use a large clearance antipad in ground planes to reduce capacitance. These concepts are depicted in Figure 11.

Figure 11. This is an edge view of a slice through a circuit board, depicting preferred IC routing for a receiver.



Unused pads at vias should be removed to reduce capacitance. A pad is unused if it is not connected to a trace, see Figure 12. Test pads should not be used.

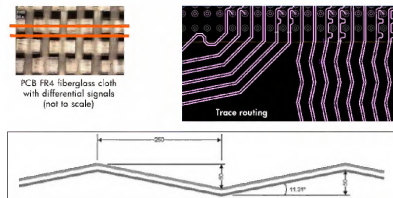
Figure 12. Remove non-functional pads



Using data from a fiber weave investigation HP determined that the differences found among PCB fiber weaves were large enough at 10Gb/s to affect signal integrity. For designs that expect to operate at speeds of 5Gbps or greater, traces should not be routed parallel to the rectangular edges

of a board for more than two inches of cumulative distance. The length of trace segments that are parallel to the edge of a board should be minimized as much as possible. Traces should be at an angle of at least 10° to a board edge. This is to reduce the effects of PCB fiber weave on board impedance and propagation velocity on differential traces (see Figure 13).

Figure 13. Zig-Zag routing example



Summary

The BladeSystem c-Class enclosure was designed to ensure that it could support new high-speed technologies and their demand for both bandwidth and power for at least 5 to 7 years.

Signal integrity issues arise as transmission frequencies increase. These issues require specific engineering to ensure that electronic circuits communicate reliably in the presence of signal impairments. This increased complexity has created a need for development of design guidelines and created a requirement for strict adherence to the resultant guidelines. The design must leverage industry standard components and interoperate with the dozens of companies and multiple divisions involved.

HP has substantially engaged in signal integrity engineering to guarantee success and manage costs effectively. The effort modularized the system into its component parts, created and applied significant modeling/simulation techniques, and then verified, through empirical data, that the models accurately reflected the physical system. HP created and maintains an extensive documentation/specification library that with adherence yields design success and customer satisfaction.

For more information

For additional information, refer to the resources listed below.

Resource description	Web address
Industry Standard Server Technology Communications	www.hp.com/servers/technology
HP BladeSystem	www.hp.com/go/bladeSystem

Call to action

Send comments about this paper to TechCom@HP.com.

© 2009 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

GbX is a registered trademark of Amphenol Corporation.

GigArray is a registered trademark of FCI.

TC0904011B

April 2009

